A generalised solution to the String Cut Problem

Paul Turner, November 2020

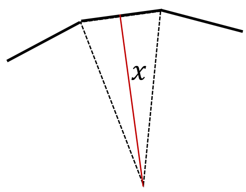
In a classroom exercise, a flexible string of length L is cut into two pieces. One piece is formed into a square and the other into a circle. Students are asked where the string should be cut to minimise the combined areas of the circle and square. Using calculus, it is shown that the minimum occurs when the perimeters are such that the circle is inscribed in the square.

On one occasion, a student had an insight that led to the correct conclusion without the need for differentiation. Years later, we looked closely at the student's idea to confirm its validity and to understand why it worked. We soon saw that not only was the idea correct in the case of the circle and square combination, but that a similar result would also hold when the parts of a string are formed into any regular polygon and a circle. That is, the minimum combined area occurs when the circle is inscribed in the polygon.

Let the fixed length of the string be L, and let c, where $0 \leq c \leq L$, be the variable portion that forms the circle. Then, the regular polygon has perimeter $p = L - c$.

The circle with circumference c has radius $r = \frac{c}{2\pi}$. Its area is $\frac{c^2}{4\pi}$.

The regular n-gon with $p = L - c$ has a distance x measured from the centre to the midpoint of each edge.



The angle subtended at the centre by an edge is $\frac{2\pi}{n}$. Thus, we have $x = \frac{\frac{p}{2n}}{\tan\frac{\pi}{n}}$ and the area of the polygon is $\frac{p^2}{4n\tan\frac{\pi}{n}}$.

We verify, first, that the minimum combined area of the two shapes occurs for some non-zero value of c. A regular n-gon, $n \geq 3$, formed with the full string length, has area $\frac{L^2}{4n\tan\frac{\pi}{n}}$, and this is assumed to be less than the area of the circle formed from the same length. That is, $\frac{L^2}{4n\tan\frac{\pi}{n}} < \frac{L^2}{4\pi}$. It follows that, $n\tan\frac{\pi}{n} > \pi$.

Suppose the string is cut into equal parts of length $\frac{L}{2}$. Then, the resulting circle has area $\frac{L^2}{16\pi}$, while the polygon has area $\frac{L^2}{16n\tan\frac{\pi}{n}}$. The combination of the two areas is $\frac{L^2}{16}\cdot\left(\frac{1}{\pi}+\frac{1}{n\tan\frac{\pi}{n}}\right)<\frac{L^2}{16}\cdot\frac{2}{n\tan\frac{\pi}{n}}$, which is clearly smaller than the area of the polygon alone. Hence, the minimum combined area must be attained not when $c=0$ but when the string is cut somewhere between its endpoints.

The two string-lengths c and $L-c$, correspond to the perimeters of the circle and n-gon, respectively. If either is changed by a small amount δ, the other is changed by $-\delta$, and there are consequent small changes in the respective areas.

The critical observation is that the minimum combined area is reached when a small increase in one area is precisely matched by the resulting decrease in the other. If this were not so, it would be possible to attain a smaller total by moving the cut in one direction or the other.

Thus, when $c\to c+\delta$, the area of the polygon changes by $\frac{\left(L-(c+\delta)\right)^2}{4n\tan\frac{\pi}{n}}-\frac{(L-c)^2}{4n\tan\frac{\pi}{n}}=\frac{-2(L-c)\delta+\delta^2}{4n\tan\frac{\pi}{n}}$ and the area of the circle changes by $\frac{(c+\delta)^2}{4\pi}-\frac{c^2}{4\pi}=\frac{2c\delta+\delta^2}{4\pi}$.

We require these changes to sum to zero as $\delta\to 0$. Therefore, $\frac{2c\delta+\delta^2}{4\pi}=\frac{2(L-c)\delta+\delta^2}{4n\tan\frac{\pi}{n}}$ and thus, in the limit, $cn\tan\frac{\pi}{n}=\pi(L-c)$.

Referring to the earlier definitions, we have $c=2\pi r$ and $L-c=p=2n\,x\tan\frac{\pi}{n}$. Thus, on substituting for these quantities, it follows that $r=x$.

We conclude that the minimum combined area is attained when the circle is inscribed in the regular n-gon for all $n\geq 3$.

An abstraction

By forgetting that we are dealing with string and with the areas of shapes, the idea can be presented more abstractly, with the possibility of further applications.

Consider a parameter $t\in\mathbb{R}^+\cup\{0\}$ that can vary so that $0\leq t\leq T$. Two continuous functions f and g, both increasing or both decreasing, are formed together with the sum $h(t)=f(t)+g(T-t)$. The function h attains its minimum value either at an endpoint or somewhere between 0 and T.

As before, if the critical value of t is in the open interval $0<t<T$, then at the minimum, a small change in t leads to a change in f that will be equal and opposite to the change in g. That is, $\lim_{\delta\to 0}f(t+\delta)-f(t)+g\big(T-(t+\delta)\big)-g(T-t)=0$.

Thus, given the form of the functions f and g, it will be possible to find the critical value of the parameter t.